



F U N D A Ç Ã O
GETULIO VARGAS

EPGE

Escola de Pós-Graduação
em Economia

Ensaio Econômico

Escola de

Pós-Graduação

em Economia

da Fundação

Getúlio Vargas

Nº 728

ISSN 0104-8910

A Common-Feature Approach for Testing Present-Value Restrictions with Financial Data

Hecq, Alain, Issler, João Victor

Fevereiro de 2012

URL: <http://hdl.handle.net/10438/9303>

Os artigos publicados são de inteira responsabilidade de seus autores. As opiniões neles emitidas não exprimem, necessariamente, o ponto de vista da Fundação Getúlio Vargas.

ESCOLA DE PÓS-GRADUAÇÃO EM ECONOMIA

Diretor Geral: Rubens Penha Cysne

Vice-Diretor: Aloisio Araujo

Diretor de Ensino: Carlos Eugênio da Costa

Diretor de Pesquisa: Luis Henrique Bertolino Braidó

Direção de Controle e Planejamento: Humberto Moreira

Vice-Diretor de Graduação: André Arruda Villela

Alain, Hecq,

A Common-Feature Approach for Testing Present-Value
Restrictions with Financial Data/ Hecq, Alain, Issler,
João Victor - Rio de Janeiro : FGV,EPGE, 2012
31p. - (Ensaio Econômico; 728)

Inclui bibliografia.

CDD-330

A Common-Feature Approach for Testing Present-Value Restrictions with Financial Data

Alain Hecq
Maastricht University

João Victor Issler*
Getulio Vargas Foundation

February 7, 2012

Abstract

It is well known that cointegration between the level of two variables (labeled Y_t and y_t in this paper) is a necessary condition to assess the empirical validity of a present-value model (PV and PVM, respectively, hereafter) linking them. The work on cointegration has been so prevalent that it is often overlooked that another necessary condition for the PVM to hold is that the forecast error entailed by the model is orthogonal to the past. The basis of this result is the use of *rational expectations* in forecasting future values of variables in the PVM. If this condition fails, the present-value equation will not be valid, since it will contain an additional term capturing the (non-zero) conditional expected value of future error terms.

Our article has a few novel contributions, but two stand out. First, in testing for PVMs, we advise to split the restrictions implied by PV relationships into orthogonality conditions (or reduced rank restrictions) before additional tests on the value of parameters. We show that PV relationships entail a weak-form common feature relationship as in Hecq, Palm, and Urbain (2006) and in Athanasopoulos, Guillén, Issler and Vahid (2011) and also a polynomial serial-correlation common feature relationship as in Cubadda and Hecq (2001), which represent restrictions on dynamic models which allow several tests for the existence of PV relationships to be used. Because these relationships occur mostly with financial data, we propose tests based on generalized method of moment (GMM) estimates, where it is straightforward to propose robust tests in the presence of heteroskedasticity. We also propose a robust Wald test developed to investigate the presence of reduced rank models. Their performance is evaluated in a Monte-Carlo exercise.

*Corresponding address: João Victor Issler, Graduate School of Economics - EPGE, Getulio Vargas Foundation, Praia de Botafogo 190 s. 1100, Rio de Janeiro, RJ 22250-900, Brazil. E-mail: Joao.Issler@fgv.br. The authors gratefully acknowledge the financial support of INCT-Brazil (CNPq and FAPERJ). We also thank Søren Johansen for bringing to our attention several of his recent papers.

Second, in the context of asset pricing, we propose applying a permanent-transitory (PT) decomposition based on Beveridge and Nelson (1981), which focus on extracting the long-run component of asset prices, a key concept in modern financial theory as discussed in Alvarez and Jermann (2005), Hansen and Scheinkman (2009), and Nieuwerburgh, Lustig, Verdelhan (2010). Here again we can exploit the results developed in the common cycle literature to easily extract permanent and transitory components under both long and also short-run restrictions.

The techniques discussed herein are applied to long span annual data on long- and short-term interest rates and on price and dividend for the U.S. economy. In both applications we do not reject the existence of a common cyclical feature vector linking these two series. Extracting the long-run component shows the usefulness of our approach and highlights the presence of asset-pricing bubbles.

JEL: C22, C32

Keywords: present value, common cycles, cointegration, interest rates, prices and dividends.

1 Introduction

Since Campbell and Shiller (1987), it is well known that cointegration between the level of two variables (labeled Y_t and y_t in this paper) is a necessary condition to assess the empirical validity of a present-value model (PV and PVM, respectively, hereafter) linking them. To make Y_t and y_t concrete, in our context, we consider a long-run relationship between prices and dividends, long and short-term interest rates or between consumption and income. If they are integrated processes, they will cointegrate; see also Campbell (1987) and Campbell and Deaton (1989), *inter alia*, which are reviewed in Engsted (2002), and the interesting recent contribution of Johansen and Swensen (2011).

The work on cointegration has been so prevalent that it is often overlooked that another necessary condition for the PVM to hold is that the forecast error entailed by the model is orthogonal to the past. We can refer to Hansen and Sargent (1981, 1991) and Baillie (1989) for initial work on rational expectations linked to PVMs, and Johansen and Swensen (1999, 2004, 2011) and Johansen (2000) for a recent fresh look on the subject. The basis of this result is the use of *rational expectations* in forecasting future values of variables in the PVM. Indeed, as shown by Campbell in a discussion on “saving for a rainy day”, there is a first-order stochastic difference equation generating the PVM relating saving and the expected discounted value of all future income changes, where its error term must be unforecastable regarding past information, i.e., must have a zero conditional expectation. If this condition fails, the present-value equation will not be valid, since it will contain an additional term capturing the (non-zero) conditional expected value of future error terms.

On the one hand, regarding the first-order stochastic difference equation generating PVMs,

cointegration imposes the transversality condition allowing to discard the limit $I(0)$ combination of Y_t and y_t . On the other hand, the existence of an unforecastable linear combination of the $I(0)$ series in the difference equation generating the PVM is crucial to guarantee that the dynamic behavior of the variables in the PVM is consistent with theory. We need both conditions to validate PVMs. Thus, it is ideal to work with an integrated econometric framework encompassing the joint existence of these two phenomena.

This is the starting point this article. We show that the orthogonality conditions entailed by PVMs are equivalent to reduced rank restrictions for dynamic systems, which imply the presence of common cyclical features¹ for different econometric representations containing them. Since these restrictions apply to the short-run behavior of the variables in PVMs, they complement the long-run restrictions implied by cointegration. Because the *toolkit* of the common-cycle literature allows the joint treatment of these two types of restrictions in dynamic models (usually a vector autoregression (VAR) model or a VECM, but not restricted to them), it is ideally designed to be the basis of the investigation of PVMs.

Given the well known problem that existing tests appear to reject PV theory too often, even when theory looks appropriate, we propose first to look at the statistical properties of the data in terms of cointegration and common-cyclical features, later applying additional tests on specific values of the parameters of the PVM. Testing every restriction implied by the PVM in “one shot” makes it difficult to interpret possible rejections of the joint hypotheses underlying the model. This is an important issue, since, as shown below, a small modification in the timing of the PV equation changes the cross-equation restrictions and hence the value of some parameters but neither the orthogonality condition embedded on PVMs nor the reduced-rank properties of the cointegrated VAR are affected by these changes. Hence, reduced-rank tests should be preferred to full cross-equation restriction tests.

Focussing on the dynamics of PVMs, this paper has two main contributions. First, we show that PV relationships entail a weak-form common feature restriction as in Hecq *et al.* (2006) and in Athanasopoulos *et al.* (2011) in the vector error correction representation for Y_t and y_t as well as a polynomial serial correlation common feature relationship (see Cubadda and Hecq, 2001) in the VAR representation for Δy_t and the cointegrating relationship $Y_t - \theta y_t$. These represent restrictions on the short-run dynamics of VECM/VAR models. Taken together with the long-run restriction implied by cointegration, we are able to devise new tests for the existence of PV relationships. Because PVMs occur mostly with financial data, the tests proposed here are robust to the presence of heteroskedasticity of unknown form. Their good performance is confirmed in a Monte-carlo exercise, where their empirical size is investigated.

¹See, *inter alia*, Engle and Kozicki (1993), Vahid and Engle (1993, 1997), Cubadda and Hecq (2001), Vahid and Issler (2002), Hecq, Palm, and Urbain (2006), Athanasopoulos, Guillén, Issler and Vahid (2011).

Second, in the context of asset pricing, we propose employing a permanent-transitory (PT) decomposition based on Beveridge and Nelson (1981), which focus on extracting the long-run component of asset prices, a key concept in modern financial theory as discussed in Alvarez and Jermann (2005), Hansen and Scheinkman (2009), and Nieuwerburgh, Lustig, and Verdelhan (2010). Here, we advance with respect to standard Beveridge and Nelson (BN) decompositions in which we compute the transitory component by directly using the PVM short-run restrictions under common-cyclical features. Since our decomposition is in the Beveridge and Nelson class, the permanent component is a fairly good approximation of the limit to which the conditional expectation of asset prices converges to. Thus, we can think of asset-price deviations from trend as bubbles in asset prices², for which there has been a renewed interest since the last global recession.

The techniques discussed in this paper are applied to two different data sets. The first contains annual long- and short-term interest rates for the U.S., ranging from 1871 to 2011. The second application involves price and dividends for the S&P Composite index on the period 1871-2010. In both case we test the degree of integration, the presence of cointegration and common cycles. We also extract the common long-run component using our proposed PT decomposition. The results show promise for its application since we were able to associate peaks and troughs in our bubble estimate with peaks and troughs of the stock market.

The rest of the paper is divided as follows. Section 2 reviews PV formulas and notations. Section 3 discusses the types of restrictions a simple present value model imply for the VECM as well as for a transformed VAR. Section 4 discusses different tests of PVMs, where their small-sample performance is evaluated in Section 5. Section 6 discusses the novel permanent transitory decomposition under PVMs which is used to measure asset pricing bubbles in the empirical part in Section 7. Finally, Section 8 concludes.

2 A present value equation

Consider the present value equation³:

$$Y_t = \theta(1 - \delta) \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t y_{t+i}, \quad (1)$$

which states that Y_t is a linear function of the present discounted value of expected future y_t , where $\mathbb{E}_t(\cdot)$ is the conditional expectation operator, using information up to t as the information

²In principle, since the limit conditional expectation does not depend on the stochastic properties of asset prices, i.e., on whether asset prices have or not a unit root, bubbles can be measured regardless of one's belief in unit roots, although in practice one usually takes a stand on the issue.

³For simplicity, we do not include a constant term c at this level of presentation as some papers do.

set. In most cases Y_t and y_t are $I(1)$ variables. Examples of Y_t and y_t include, respectively: long and short-term interest rates, stock prices and dividends, personal consumption and disposable income, etc. (see the survey Engsted, 2002). Here and elsewhere, it is assumed constant expected returns with a discount factor $\delta = \frac{1}{1+r}$. The coefficient θ is a factor of proportionality. For example, $\theta = \delta/(1 - \delta)$ in the price-dividend relationship; $\theta = 1$ for the interest rates case and the link with the discount factor is given by the term structure of the interest rates (see *inter alia* Campbell and Shiller, 1987; Chow, 1984; Johansen and Swensen, 2011). The choice of θ only impacts the value of the cointegrating vector. Hence, here, in what follows, we set its value equal to $\theta = \delta/(1 - \delta)$, such that:

$$Y_t = \delta \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t y_{t+i}. \quad (2)$$

Following Campbell and Shiller (1987), the actual spread is defined as:

$$S_t \equiv Y_t - \frac{\delta}{1 - \delta} y_t, \quad (3)$$

where S_t is $I(0)$ if Y_t and y_t are cointegrated. Subtracting $\frac{\delta}{1 - \delta} y_t$ from both sides of (2) produces the theoretical spread S'_t :

$$S'_t = \frac{\delta}{1 - \delta} \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t \Delta y_{t+i}. \quad (4)$$

This shows that series must be theoretically cointegrated because the right-hand side is a function of $I(0)$ terms with exponentially decreasing weights. Further, subtracting $\delta \mathbb{E}_t Y_{t+1} = \delta \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t y_{t+i+1}$ from Y_t in (4), we obtain:

$$Y_t = \delta \mathbb{E}_t Y_{t+1} + \delta y_t. \quad (5)$$

From (5), if one adds and subtracts δY_t , leading to $Y_t = \delta \mathbb{E}_t \Delta Y_{t+1} + \delta Y_t + \delta y_t$ or $(1 - \delta) Y_t = \delta \mathbb{E}_t \Delta Y_{t+1} + \delta y_t$, one finally obtains:

$$S''_t = \frac{\delta}{1 - \delta} \mathbb{E}_t \Delta Y_{t+1}. \quad (6)$$

Equation (6) gives the spread as a function of one-step ahead forecasts of ΔY_{t+1} .

We can always perform the following decomposition:

$$\Delta Y_{t+1} = \mathbb{E}_t \Delta Y_{t+1} + \underbrace{(\Delta Y_{t+1} - \mathbb{E}_t \Delta Y_{t+1})}_{u_{t+1}}. \quad (7)$$

Plugging (7) into (6), and lagging the whole equation by one period we have:

$$S_{t-1} = \frac{\delta}{1-\delta} \Delta Y_t + u_t \quad (8)$$

or alternatively,

$$\Delta Y_t = \frac{1-\delta}{\delta} S_{t-1} + v_t \quad (9)$$

where u_t (or $v_t = -\frac{1-\delta}{\delta} u_t$) is orthogonal to the past in expectation. From (9) we also obtain:

$$\begin{aligned} (1-\delta)S_{t-1} &= \delta \Delta Y_t + (1-\delta)u_t \\ S_{t-1} - \delta Y_{t-1} + \delta \frac{\delta}{1-\delta} y_{t-1} &= \delta Y_t - \delta Y_{t-1} + \left\{ \delta \frac{\delta}{1-\delta} y_t - \delta \frac{\delta}{1-\delta} y_t \right\} + (1-\delta)u_t \\ S_{t-1} &= \delta S_t + \delta \frac{\delta}{1-\delta} \Delta y_t + (1-\delta)u_t \end{aligned} \quad (10)$$

which gives

$$S_t = \frac{1}{\delta} S_{t-1} - \frac{\delta}{1-\delta} \Delta y_t + \varepsilon_t \quad (11)$$

with $\varepsilon_t = \frac{(1-\delta)}{\delta} u_t$.

As stressed by Campbell (1987), in the context of saving, equation (11) plays a very important role: it is the first order stochastic difference equation that generates the PVM. There are two important conditions to go from (11) to (4): cointegration delivers the transversality condition $\lim_{k \rightarrow \infty} \delta^k \mathbb{E}_t(S_{t+k}) = 0$, whereas unforecastability of ε_t regarding the past, i.e., $\mathbb{E}_t(\varepsilon_{t+1}) = 0$, ensures that there is no additional term in the right-hand side of (4) invalidating it. The first represents a long-run restriction between Y_t and y_t . The second restricts the dynamics of the stationary representation of the system, making S_t and Δy_t specific functions of their own past alone. Thus, they can be viewed as short-run restrictions on the behavior of S_t and Δy_t . These are exactly the types of restrictions studied in the common cycle literature. Therefore, applying the *toolkit* developed there allows a fresh view of PVMs as we show below.

Johansen and Swensen (2011) discuss the properties of the three spreads S_t , S'_t , and S''_t . Their setup is slightly different than ours, since, in (1), they define the present-value relationship to be $Y_t = \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t y_{t+i}$ instead of $Y_t = \theta(1-\delta) \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t y_{t+i}$, i.e., they discount only future values of y_t and not its current value. Some authors prefer the latter to the former using the argument that, in the discrete time setup, the cash flow is accrued at the end of every period. Here, we follow Campbell and Shiller in their choice of PV formula using (1). The cointegrating vector is not affected by this choice, but the short-run dynamic coefficients are, as we shall see in the next section. For that reason, some of our results are not identical to those in Johansen and Swensen

(2011).

3 Common cyclical feature restrictions: Model representations

Assume that the bivariate system for the $I(1)$ series $(Y_t, y_t)'$ follows a $VAR(p)$ in levels, and that $S_t = Y_t - \theta y_t$ is the stationary error-correction term. In the price-dividend case $\theta = \frac{\delta}{1-\delta}$. The corresponding vector error-correction model (VECM) representation is given by:

$$\begin{pmatrix} \Delta Y_t \\ \Delta y_t \end{pmatrix} = \Gamma_1 \begin{pmatrix} \Delta Y_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \dots + \Gamma_{p-1} \begin{pmatrix} \Delta Y_{t-p+1} \\ \Delta y_{t-p+1} \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} S_{t-1} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} \quad (12)$$

where we assume that the disturbance terms are white noise and that conditions for avoiding $I(2)$ -ness are met. The Γ_i s are the short-run coefficient matrices, and α_1 and α_2 are the loadings on the error-correcting term.

As is well known, PV relationships imply restrictions on dynamic models of the data. Campbell and Shiller (1987) and others have exploited the fact that VARs have cross-equation restrictions. Here, however, we exploit a different nature of these restrictions – the fact that there are also reduced-rank restrictions for the VECM (12), which opens up the application of the common cyclical feature *toolkit* in dealing with PVMs.

Proposition 1 *If the elements of $(Y_t, y_t)'$ obey a PV relationship as in (8), i.e., $S_{t-1} = \frac{\delta}{1-\delta} \Delta Y_t + u_t$, then, their VECM obeys a weak-form common feature relationship (see Hecq et al., 2006, and Athanasopoulos et al., 2011): there exists a 1×2 vector γ' such that $\gamma' \Gamma_1 = \gamma' \Gamma_2 = \dots = \gamma' \Gamma_{p-1} = 0$, but $\gamma' \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \neq 0$. Moreover, $\gamma' = (1 : 0)$, the first row of every Γ_i , $i = 1 \dots p-1$, must be zero, and the following restriction must also be met: $\alpha_1 = \frac{1-\delta}{\delta}$.*

The usual cross-equation restriction within the VAR and proposed by Campbell and Shiller (1987) can also be seen from a transformed VAR on S_t and Δy_t ; see Johansen and Swensen (2011).

To go from the VECM (12) to the transformed VAR representation we use $C = \begin{bmatrix} 1 & 0 \\ -\theta & 1 \end{bmatrix}'$, the 2×2 nonsingular matrix formed by stacking the transpose of the cointegrating vector $\begin{bmatrix} 1 \\ -\theta \end{bmatrix}$ and

the transpose of the selection vector $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, such that $C \begin{pmatrix} \Delta Y_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} \Delta S_t \\ \Delta y_t \end{pmatrix}$. Premultiplying

both sides of (12) by C , and solving for S_t and Δy_t , we obtain:

$$\begin{pmatrix} S_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} \underline{\Gamma}_{11}(L) & \underline{\Gamma}_{12}(L) \\ \underline{\Gamma}_{21}(L) & \underline{\Gamma}_{22}(L) \end{pmatrix} \begin{pmatrix} S_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \quad (13)$$

where $\underline{\Gamma}_{11}(L)$ and $\underline{\Gamma}_{21}(L)$ are polynomials of order $p-1$ and $\underline{\Gamma}_{12}(L)$ and $\underline{\Gamma}_{22}(L)$ are polynomials of order $p-2$. Indeed, one important issue is to note is that the transformed VAR (13) is a VAR of order p both in S_t and in Δy_t in which the two coefficients of Δy_{t-p} are zero. Cross-equation restrictions for the system are imposed on the coefficient matrices of $\underline{\Gamma}(L) = \begin{pmatrix} \underline{\Gamma}_{11}(L) & \underline{\Gamma}_{12}(L) \\ \underline{\Gamma}_{21}(L) & \underline{\Gamma}_{22}(L) \end{pmatrix} = \underline{\Gamma}_1 + \underline{\Gamma}_2 L + \underline{\Gamma}_p L^{p-1}$ in (13).

We have the following proposition.

Proposition 2 *A PVM as in (11), i.e., $S_t = \frac{1}{\delta} S_{t-1} - \frac{\delta}{1-\delta} \Delta y_t + \varepsilon_t$, implies a polynomial serial-correlation common feature relationship (see Cubadda and Hecq, 2001) for the transformed VAR (13): there exists a vector $\tilde{\gamma}'_0$ such that $\tilde{\gamma}'_0 \underline{\Gamma}_2 = \dots = \tilde{\gamma}'_0 \underline{\Gamma}_p = 0$, with $\tilde{\gamma}'_0 \underline{\Gamma}_1 = \tilde{\gamma}'_1 \neq 0$. Moreover in the PVM $\tilde{\gamma}'_0 = (1 : \frac{\delta}{1-\delta})$ and $\tilde{\gamma}'_1 = (-\frac{1}{\delta} : 0)$.*

Thus, a PVM entails cointegration and additional orthogonality conditions associated with reduced rank restrictions in VECMs or transformed VARs. It is interesting to gain some intuition on this result. For that, we resort to the triangular representation of cointegrated systems used, *inter alia*, by Phillips and Hansen (1990) and Phillips and Loretan (1991), which was adapted to account for reduced-rank dynamics (weak-form SCCF) by Athanasopoulos *et al.* (2011). In our context, their representation for $(Y_t, y_t)'$ would be:

$$Y_t = \theta y_t + \mu_{1t} \quad (14)$$

$$\Delta y_t = \mu_{2t}, \quad (15)$$

where the error terms, stacked on a vector $(\mu_{1t}, \mu_{2t})'$ follow a stationary and ergodic $VAR(p-1)$, where the coefficient matrices (2×2 matrices) have all rank one. Notice that the long-run value for Y_t is θy_t , making μ_{1t} to be the gap between the two.

Here, the reduced-rank nature of the VAR for $(\mu_{1t}, \mu_{2t})'$ is what generates common features for $(S_t, \Delta y_t)'$. To see it, subtract θy_t from both sides of (14), to get:

$$\begin{aligned} Y_t - \theta y_t &= S_t = \mu_{1t}, \\ \Delta y_t &= \mu_{2t}, \end{aligned}$$

thus, $(S_t, \Delta y_t)' = (\mu_{1t}, \mu_{2t})'$, and have the dynamic structure of equation (13), discussed in Propo-

sition 2.

One of the explanations for observing a rejection of the PVMs is the use of cross-equation restrictions that impose both reduced-rank restrictions and particular values on the parameters. Misspecifications such as proxy variables or measurement errors can affect the value of the parameters, leaving unaffected the reduced-rank restrictions. As an example, instead of the PV representation given in our Section 2, i.e. $Y_t = \theta(1 - \delta) \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t y_{t+i}$ one can find in the literature that the series Y_t is a function of the future discounted expected value of y_t such that $Y_t = \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t y_{t+i}$. Johansen and Swensen (2011) as well as Campbell, Lo and Mackinlay (1996) use that formulation when they consider the stock price at the end of the period. This slight change is not innocuous as we show next.

To see that, apply the algebra of Section 2 to $Y_t = \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t y_{t+i}$ to obtain the following expressions:

$$\Delta Y_t = -\Delta y_t + \frac{1 - \delta}{\delta} S_{t-1} + u_t, \quad (16)$$

$$\text{where } \gamma' = (1 : 1) \text{ and } \alpha_1 = \frac{1 - \delta}{\delta} \text{ in Proposition 1.} \quad (17)$$

$$S_t = -\frac{1}{(1 - \delta)} \Delta y_t + \frac{1}{\delta} S_{t-1} + v_t, \quad (18)$$

$$\text{where } \tilde{\gamma}'_0 = (1 : \frac{1}{1 - \delta}) \text{ and } \tilde{\gamma}'_1 = (-\frac{1}{\delta} : 0) \text{ in Proposition 2.} \quad (19)$$

What emerges now is that the unpredictable linear combinations involve three variables: ΔY_t , Δy_t , and S_t , both in the VECM and the transformed VAR. Moreover the values of the parameters are now different from before – the weights used in the linear combinations (16) and (18) differ from the ones in (8) and (11), respectively.

Put differently, regarding the use of $Y_t = \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t y_{t+i}$ versus $Y_t = \theta(1 - \delta) \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t y_{t+i}$, respectively, yields the following orthogonality conditions for each specific difference equation:

$$\mathbb{E}_{t-1} \left[\Delta Y_t + \Delta y_t - \frac{1 - \delta}{\delta} S_{t-1} \right] = 0, \text{ versus } \mathbb{E}_{t-1} \left[\Delta Y_t - \frac{1 - \delta}{\delta} S_{t-1} \right] = 0, \text{ and,} \quad (20)$$

$$\mathbb{E}_{t-1} \left[S_t + \frac{1}{(1 - \delta)} \Delta y_t - \frac{1}{\delta} S_{t-1} \right] = 0, \text{ versus } \mathbb{E}_{t-1} \left[S_t + \frac{\delta}{1 - \delta} \Delta y_t - \frac{1}{\delta} S_{t-1} \right] = 0. \quad (21)$$

Despite the differences in parameter values in the linear combinations above, the existence of a reduced-rank model is not affected by how one writes the PV equation linking Y_t and y_t . Hence, the reduced-rank properties of the VECM and of the transformed VAR are invariant to this choice: in both cases, there exists weak-form SCCF for the VECM and the transformed VAR.

4 Testing present-value models: a for common-cycle approach

The discussion in the last section suggests that, for integrated Y_t and Δy_t , there are three different instances in which we can investigate the validity of PVMs. First, the cointegration test for Y_t and y_t , if both are $I(1)$. Second, the (invariant) rank restrictions in the VECM or the transformed VAR. Third, the coefficient restrictions and unpredictability properties for linear combinations in (20) and (21).

In order to test for PVMs, we propose the following steps:

1. Choose consistently the order of the $VAR(p)$ for the joint $I(1)$ process $(Y_t, y_t)'$ using different information criteria. Alternatively, we can compute a robust Wald test for the null hypothesis that the last coefficient matrix in the VAR has zero coefficients (see the empirical section).
2. Given our choice of p , test for the existence of cointegration between Y_t and y_t . If that is the case (there exists one cointegrating vector), estimate the long-run coefficient θ , in $S_t = Y_t - \theta y_t$, super-consistently using the likelihood-based trace test proposed by Johansen (1995). Alternatively, the Engle and Granger (1987) regression test can be carried out. In either case, form $\hat{S}_t = Y_t - \hat{\theta} y_t$. If there is no cointegration, the PVM is rejected.
3. Given p and \hat{S}_t , test for the weak form common feature using a reduced rank test for $(\Delta Y_t, \Delta y_t)'$. We present in this section both multivariate approaches (e.g. a canonical correlation analysis) and a single-equation approach (e.g. GMM). Because most present-value relationships apply to heteroskedastic financial data, one may prefer a GMM framework on the basis that it easily embeds robust variance-covariance matrices for parameters estimates. Indeed the canonical correlation approach assumes *i.i.d.* disturbances. However, we also provide a multivariate robust Wald test to investigate the reduced rank hypothesis under GARCH innovations.

Note that we can improve over steps 1 to 3 using steps 4 and/or 5 below. Given that we only work with bivariate systems for a relatively large number of observations in this paper we do not introduce those small sample improvements into our analysis. But, these are:

4. Integrate steps 2 and 3, estimating jointly long-run and short-run parameters as in Centoni, Cubadda and Hecq (2008).
5. Integrate steps 1, 2 and 3, estimating jointly the lag length of the VAR and long-run and short-run parameters as in Athanasopoulos *et al.* (2011).

4.1 Multivariate tests

4.1.1 LR tests for i.i.d. disturbances

The canonical-correlation approach entails the use of a likelihood ratio (reduced-rank regression) test for the weak-form common features in the $VECM(p-1)$ for $(\Delta Y_t, \Delta y_t)'$. It can be undertaken using the canonical-correlation test on zero eigenvalues, which are computed from:

$$CanCor \left\{ \begin{pmatrix} \Delta Y_t \\ \Delta y_t \end{pmatrix}, \begin{pmatrix} \Delta Y_{t-1} \\ \vdots \\ \Delta Y_{t-p+1} \\ \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \end{pmatrix} \mid (D_t, \hat{S}_{t-1}) \right\}, \quad (22)$$

where $CanCor \{X_t, W_t | G_t\}$ denotes the computation of canonical correlations between the two sets of variables X_t and W_t , concentrating out the effect of G_t (deterministic terms and a disequilibrium error-correction term) by multivariate least squares. The previous program (22) is numerically equivalent to

$$CanCor \left\{ \begin{pmatrix} \Delta Y_t \\ \Delta y_t \\ \hat{S}_{t-1} \end{pmatrix}, \begin{pmatrix} \Delta Y_{t-1} \\ \vdots \\ \Delta Y_{t-p+1} \\ \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \\ \hat{S}_{t-1} \end{pmatrix} \mid D_t \right\} \quad (23)$$

which is more convenient to directly obtain the coefficient of \hat{S}_{t-1} in (11). The likelihood ratio test, denoted by ξ_{LR} , considers the null hypothesis that there exist at least s common feature vectors. It is obtained in

$$\xi_{LR} = -T \sum_{i=1}^s \ln(1 - \hat{\lambda}_i), \quad s = 1, 2, \quad (24)$$

where $\hat{\lambda}_i$ are the i -th smallest squared canonical correlations computed from (22) or (23) above, namely from

$$\hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XW} \hat{\Sigma}_{WW}^{-1} \hat{\Sigma}_{WX}, \quad (25)$$

or similarly from the symmetric matrix

$$\hat{\Sigma}_{XX}^{-1/2} \hat{\Sigma}_{XW} \hat{\Sigma}_{WW}^{-1} \hat{\Sigma}_{WX} \hat{\Sigma}_{XX}^{-1/2}, \quad (26)$$

where $\hat{\Sigma}_{ij}$ are the empirical covariance matrices, $i, j = X, W$.

In the bivariate case, the unrestricted VECM has $4(p-1)+2$ parameters, whereas the restricted model has $2(p-1)+2+1$. The number of restrictions when testing the hypothesis that there exists one WF common feature is then $2(p-1)-1 = 2p-3$ for $p > 1$.⁴ As proposed in Issler and Vahid (2001), we can obtain the same statistics by computing twice the difference between the log-likelihood in the unrestricted VECM ($p-1$) for $(\Delta Y_t, \Delta y_t)'$ and in the pseudo-structural form estimated by FIML:

$$\begin{pmatrix} 1 & -\gamma_0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta Y_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} 0 \\ \tilde{\Gamma}_1 \end{pmatrix} \begin{pmatrix} \Delta Y_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \dots + \begin{pmatrix} 0 \\ \tilde{\Gamma}_{p-1} \end{pmatrix} \begin{pmatrix} \Delta Y_{t-p+1} \\ \Delta y_{t-p+1} \end{pmatrix} + \begin{pmatrix} (\alpha_1 - \gamma_0 \alpha_2) \\ \tilde{\alpha}_2 \end{pmatrix} S_{t-1} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}.$$

For the transformed VAR the restriction underlying the restricted PSCCF can be tested using:

$$CanCor \left\{ \begin{pmatrix} \hat{S}_t \\ \Delta y_t \\ \hat{S}_{t-1} \end{pmatrix}, \begin{pmatrix} \hat{S}_{t-1} \\ \hat{S}_{t-2} \\ \vdots \\ \hat{S}_{t-p} \\ \Delta y_{t-2} \\ \vdots \\ \Delta y_{t-p+1} \end{pmatrix} \mid D_t \right\},$$

where the number of parameters in the unrestricted model is $4(p-1)+2$; the restricted model has $4+2(p-2)+1+1$, the number of restrictions is $2p-4$ in case of unrestricted $\tilde{\gamma}_1$

$$\begin{aligned} & \begin{pmatrix} 1 & -\tilde{\gamma}_0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} S_t \\ \Delta y_t \end{pmatrix} \\ = & \begin{pmatrix} \tilde{\Gamma}_{1a} \\ \tilde{\Gamma}_{1b} \end{pmatrix} \begin{pmatrix} S_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \dots + \begin{pmatrix} 0 \\ \tilde{\Gamma}_{p-1} \end{pmatrix} \begin{pmatrix} S_{t-p+1} \\ \Delta y_{t-p+1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \tilde{\Gamma}_{p,p} & 0 \end{pmatrix} \begin{pmatrix} S_{t-p} \\ \Delta y_{t-p} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}. \end{aligned}$$

⁴In the VECM, the general formula for n series that can be annihilated by s combinations is $sn(p-1) - s(n-s)$.

If $\tilde{\gamma}_1$ is restricted we have $2p - 3$ restrictions and the pseudo structural form is

$$\begin{aligned} & \begin{pmatrix} 1 & -\tilde{\gamma}_0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} S_t \\ \Delta y_t \end{pmatrix} + \begin{pmatrix} \vartheta_1 & 0 \\ \tilde{\Gamma}_{2,1} & \tilde{\Gamma}_{2,2} \end{pmatrix} \begin{pmatrix} S_{t-1} \\ \Delta y_{t-1} \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ \tilde{\Gamma}_{p-1} \end{pmatrix} \begin{pmatrix} S_{t-2} \\ \Delta y_{t-2} \end{pmatrix} + \cdots + \begin{pmatrix} 0 & 0 \\ \tilde{\Gamma}_{p,p} & 0 \end{pmatrix} \begin{pmatrix} S_{t-p} \\ \Delta y_{t-p} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \end{aligned}$$

Notice that this set of rank restrictions are identical to the ones in Campbell and Shiller (1987) if one imposes zero restrictions in the last matrix coefficient in their setup⁵. Campbell and Shiller also take into account the fact that there are further restrictions on the parameters coming from the economic theory. Thus, the rank condition is just a *necessary condition* for the PVM to hold, but there are additional restrictions on matrices coefficients that have to be met for PV theory to be correct.

The proposed approach to testing PVMs here is to first test the rank condition (necessary) without imposing yet any further parameter restrictions. As argued above, the rank condition is invariant to how we write the PV equation linking Y_t and y_t . If not rejected, then we can test the additional restrictions on matrices coefficients, which are not invariant to how we write the PV equation. Putting more weight on invariant restrictions satisfies robustness, since, not only a different definition of the timing of Y_t and/or y_t , but also the presence of measurement error, data revisions, all will lead to the correct rank condition to be met but imply different parameter values in the difference equation generating PVMs.

An additional reason to follow this path is that we will be able to split both effects, shedding light on the exact reason for rejecting theory if that is the case. Understanding why we reject a given PVM is an important issue, since different authors have complained that cross-equation restriction tests reject PVMs too often, even in cases where theory is firmly believed to hold and that graphical analysis seems to support that view.

4.1.2 A robust Wald test

Candelon *et al.* (2005) have illustrated in a Monte Carlo exercise that ξ_{LR} has large size distortions in the presence of GARCH disturbances. The solution proposed there was to use nonparametric tests or a GMM approach (see also the next subsection) in which the variance-covariance matrix is the robust HCSE variance-covariance proposed by White.

In order to find a multivariate robust counterpart to the canonical correlation approach, we propose to modify ξ_{LR} in two respects. First we use a Wald approach, denoted ξ_W , with ξ_W

⁵This is probably implicit in their analysis but it is not discussed in the paper itself.

asymptotically equivalent to ξ_{LR} (see Christensen *et al.* 2011). Then we robustify ξ_W , a test we denote ξ_W^{rob} , using the multivariate extension of the White's HCSE proposed in Ravikumar *et al.* (2000) for system of seemingly unrelated regressions.

To do so, let us define for the VECM the weak form reduced rank restrictions, as

$$Rvec(\Gamma_1 : \dots : \Gamma_{p-1})' = 0_{sd \times 1} \quad R = \gamma' \otimes I_d$$

R is $sd \times nd$, with d is the number of rows in the rectangular matrix $A = (\Gamma_1 : \dots : \Gamma_{p-1})'$, namely $d = 2(p-1)$. Using $\hat{\gamma}$ obtained by the eigenvectors of the canonical correlation (25), the Wald test is

$$\xi_W = (Rvec\hat{A})'(R \text{Var}(vec\hat{A}) R')^{-1}(Rvec\hat{A}),$$

with

$$\text{Var}(vec\hat{A}) = \hat{V} \otimes (\tilde{W}'\tilde{W})^{-1},$$

and with \hat{V} the empirical covariance matrix of the disturbance terms in the unrestricted models and \tilde{W} are the demeaned regressors. ξ_W is asymptotically equivalent to ξ_{LR} (see Christensen *et al.* 2011). Now in the presence of a time varying multivariate process we compute an estimator of $\text{Var}(vec\hat{A})$ robust to the presence of heteroskedasticity (see Ravikumar *et al.* 2000) such that

$$\xi_{W(s)}^{rob} = (R \text{vec}\hat{A})'(R \text{rob_Var}(vec\hat{A}) R')^{-1}(R \text{vec}\hat{A})$$

where

$$\text{rob_Var}(\hat{A}) = (I_n \otimes (\tilde{W}'\tilde{W})^{-1}) \left(\sum_{t=1}^T \hat{\eta}_t \hat{\eta}_t' \right) (I_n \otimes (\tilde{W}'\tilde{W})^{-1})$$

with

$$\hat{\eta}_t \hat{\eta}_t' = \hat{v}_t \hat{v}_t' \otimes \tilde{W}_{.t} \tilde{W}_{.t}'$$

where $\hat{v}_t = (\hat{v}_{1t}, \dots, \hat{v}_{Nt})'$ and $\tilde{W}_{.t} = (W_{1t}, \dots, W_{dt})$ the explanatory variables for observations t . ξ_W^{rob} is asymptotically equivalent to ξ_W and hence to ξ_{LR} . Note finally that Christensen *et al.* (2011) report some small-sample distortions in ξ_W , while Hecq *et al.* (2011) report size distortions for ξ_W^{rob} .

4.2 Regression-Based Tests

Testing with a GMM approach entails testing the common feature null hypothesis using an orthogonality condition between a combination of variables in the model $\left(\Delta Y_t, \Delta y_t, \hat{S}_{t-1} \right)'$ and the conditioning set W_t' . For example, in the context of (8), we would have the following moment

restrictions:

$$\mathbb{E}([\Delta Y_t - \gamma_1 \Delta y_t - \gamma_2 \hat{S}_{t-1}] \otimes W_t') = 0, \quad (27)$$

where we would have additionally to test H_0 : $\gamma_1 = 0$ and $\gamma_2 = \frac{1-\delta}{\delta}$ using a Wald test. Prior to that, we want to estimate γ_1 and γ_2 and test the validity of the over-identifying restrictions in (27). The use of IV type estimators and the associated orthogonality tests is straightforward in this context. Let us consider W_t the vector of instruments defined as before (an intercept is added). The GIVE estimator is simply the 2SLS or the IV estimator when the instruments are the past of the series, namely

$$\hat{\theta}_{GIVE} = (\Delta \mathbf{X}' \mathbf{W} (\mathbf{W}' \mathbf{W})^{-1} \mathbf{W}' \Delta \mathbf{X})^{-1} (\Delta \mathbf{X}' \mathbf{W} (\mathbf{W}' \mathbf{W})^{-1} \mathbf{W}' \Delta \mathbf{Y}), \quad (28)$$

with $\Delta \mathbf{X}_t = (\Delta y_t, \hat{S}_{t-1}, 1)'$. The validity of the orthogonality condition and consequently the presence of a common feature vector is obtained via an overidentification J-test (Hansen, 1982):

$$J_1(\theta) = T g_T(\theta; \cdot)' P_T^{-1} g_T(\theta; \cdot),$$

whose empirical counterpart is:

$$J_1(\theta_{IV}) = (\mathbf{u}' \tilde{\mathbf{W}})(\hat{\sigma}_u^2 \tilde{\mathbf{W}}' \tilde{\mathbf{W}})^{-1} (\tilde{\mathbf{W}}' \mathbf{u}).$$

The variance-covariance matrix of the orthogonality condition has under usual regularity properties the sample counterpart $\hat{P}_T = (1/T) \hat{\sigma}_u^2 (\tilde{\mathbf{W}}' \tilde{\mathbf{W}})$ with $u_t = \Delta Y_t - \hat{\gamma}_1 \Delta y_t - \hat{\gamma}_2 \hat{S}_{t-1}$. $\tilde{\mathbf{W}}$ is the demeaned \mathbf{W} , namely $\tilde{\mathbf{W}} = \mathbf{W} - \mathbf{i}(\mathbf{i}' \mathbf{i})^{-1} \mathbf{i}' \mathbf{W}$ (with $\mathbf{i} = (1 \dots 1)'$) because we do not want to impose that the common feature vector also annihilates the constant vector.

In this Section, so far, all the estimates and tests presented above embedded the assumption of homoskedasticity. This may be fine for macroeconomic data, such as consumption and income, but is clearly at odds with financial data. We also propose to correct for heteroskedasticity to achieve robust estimates. We implement the GIVE estimator by using the White's HCSE estimator such that (see Hamilton, 1994):

$$\hat{\theta}_{GMM} = (\Delta \mathbf{X}' \mathbf{W} (\mathbf{W}' \mathbf{B} \mathbf{W})^{-1} \mathbf{W}' \Delta \mathbf{X})^{-1} (\Delta \mathbf{X}' \mathbf{W} (\mathbf{W}' \mathbf{B} \mathbf{W})^{-1} \mathbf{W}' \Delta \mathbf{Y}), \quad (29)$$

where the only difference between $\hat{\theta}_{GMM}$ and the usual $\hat{\theta}_{GIVE}$ is the presence of an additional

matrix \mathbf{B} constructed such that

$$\mathbf{B} = \begin{pmatrix} u_1^2 & 0 & \cdots & 0 \\ 0 & u_2^2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & u_T^2 \end{pmatrix},$$

where $u_t = \Delta Y_t - \hat{\gamma}_{1IV} \Delta y_t - \hat{\gamma}_{2IV} \hat{S}_{t-1}$ are the residuals obtained under homoskedasticity using the GIVE estimation in a first step. For testing, we form the following new sequence of residuals:

$$u_t^* = \Delta Y_t - \hat{\gamma}_{1GMM} \Delta y_t - \hat{\gamma}_{2GMM} \hat{S}_{t-1},$$

and use these to compute a new J-test robust to heteroskedasticity:

$$J_2(\theta_{GMM}) = (\mathbf{u}^{*\prime} \tilde{\mathbf{W}})(\tilde{\mathbf{W}}' \mathbf{B} \tilde{\mathbf{W}})^{-1}(\tilde{\mathbf{W}}' \mathbf{u}^*). \quad (30)$$

Note that we have also implemented a Newey and West (1987) correction in constructing \mathbf{B} . In this case, \mathbf{B} is a band-diagonal matrix with q non-zero bands corresponding to the order of the MA process being considered in the Newey-West approach. This yields robust Newey-West estimates using (29) and a new J-test statistic using (30), which we label $J_3(\theta_{GMM})$. Since this correction applies to both heteroskedasticity and serial correlation in the error term, it can be viewed as an overkill.

5 Small sample properties of PVM tests

A small Monte Carlo simulation might help to advise the use of one of the tests considered in this paper. We use $T = 100, 500$ and $1,000$ observations with $10,000$ replications. Although $1,000$ data points might seem large it illustrates the asymptotic behavior of several testing strategies. In particular it is seen that the robust Wald statistics has some size distortions for $T = 100$ (see also simulations in Christensen *et al.*, 2011).

The lag length of the VAR in the data generating process is chosen to be $p = 3$. However, we estimate the model for $p = 2, 3$ and 5 . Notice that the model with $p = 2$ is misspecified. The DGP

that ensures $\gamma' = (1 : 0)$ is:

$$\begin{pmatrix} \Delta Y_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} 0.05 \\ 0.05 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0.5 & 0.2 \end{pmatrix} \begin{pmatrix} \Delta Y_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \dots \begin{pmatrix} 0 & 0 \\ -0.4 & 0.2 \end{pmatrix} \begin{pmatrix} \Delta Y_{t-2} \\ \Delta y_{t-2} \end{pmatrix} \\ + \begin{pmatrix} 1 \\ 0.75 \end{pmatrix} \begin{pmatrix} 1 & -\frac{\delta}{1-\delta} \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}.$$

We considered two types of error terms for the VECM above: in the first DGP, labelled DGP # 1 in Table1, the disturbance term is bivariate normal with a unit variance and a correlation of 0.5; in the second process, labelled DGP # 2, the disturbance terms are governed by a bivariate GARCH process with a yesterday news coefficient of 0.25, a coefficient of persistence of 0.74, and a long run variance equals 0.01. Note that yesterday's news coefficient is larger than what is usually found empirically (between 0.10 and 0.15). The theoretical coefficients in the relationship $\Delta Y_t = -\gamma_1 \Delta y_t + \gamma_2 S_{t-1} + u_t$ are $\gamma_1 = 0$ and $\gamma_2 = \frac{\delta}{1-\delta}$. Here, for simplicity, we set $\frac{\delta}{1-\delta} = 1$.

For all the tests described in the previous section, Table 1 reports the empirical rejection frequency at the 5% significance level (nominal size). In the *iid* case, the behavior of the six tests is rather similar, but the Wald strategy is oversized when p increases for small samples, and the J_3 -test (Newey-West correction) is undersized for $T = 100$. Results get much more worse in the presence of time varying conditional variances. With heteroskedastic data, the only test with proper size is J_2 -Test, the robust-White GMM test and, to a lesser extent, the robust Wald if T is large enough. The other tests have large size distortions, especially the likelihood-ratio test, the non robust Wald test, and usual GMM J_1 -test. Again, the J_3 -test (Newey-West correction) is undersized. Thus, for macroeconomic applications we can rely either on a robust GMM or on the robust Wald if in this latter case the number of lags in the unrestricted VECM is not too large.

As far as the estimation of the coefficients are concerned, we do not notice any systematic bias. For DGP # 1, the estimated coefficients for the mean value of γ_1 over 10,000 replications ranges over the different specifications on p and T from -0.001 to 0.008 in the worst case while those from the mean value of γ_2 are between 0.978 ($p = 5, T = 100$) and 0.999 . There are no major differences in the GARCH case (DGP #2) as expected. About the standard deviation over the simulations, it lies for DGP#1 between 0.021 ($p = 3, T = 500$) and 0.054 ($p = 5, T = 100$) for γ_1 and between 0.032 and 0.077 for γ_2 for the same specifications. In the presence of GARCH for the DGP # 2 these standard deviations are in general higher.

Table 1: Empirical size (nom. 5 percent) of common feature test statistic

$T = 100$							
	$VAR(p)$	ζ_{LR}	ζ_W	ζ_W^{rob}	J_1	J_2	J_3
DGP #1: iid	$p = 2$	5.74	5.92	7.05	5.53	5.46	4.68
	$p = 3$	6.23	6.97	9.99	5.53	5.11	3.03
	$p = 5$	6.68	8.90	17.3	4.75	3.88	0.25
DGP #2: GARCH	$p = 2$	12.5	12.9	8.47	12.3	5.35	3.17
	$p = 3$	15.5	16.8	13.2	14.2	4.73	1.39
	$p = 5$	18.1	21.7	23.6	14.9	3.29	0.12
$T = 500$							
	$VAR(p)$	ζ_{LR}	ζ_W	ζ_W^{rob}	J_1	J_2	J_3
DGP #1: iid	$p = 2$	4.91	4.92	5.11	4.84	4.84	4.78
	$p = 3$	5.14	5.23	5.93	5.04	5.09	4.62
	$p = 5$	5.07	5.35	6.94	4.78	4.62	3.49
DGP #2: GARCH	$p = 2$	23.3	23.4	6.42	23.3	4.96	3.83
	$p = 3$	30.7	30.9	8.38	30.4	4.5	2.51
	$p = 5$	38.3	38.9	11.8	37.8	3.93	1.07
$T = 1,000$							
	$VAR(p)$	ζ_{LR}	ζ_W	ζ_W^{rob}	J_1	J_2	J_3
DGP #1: iid	$p = 2$	4.92	4.94	5.07	4.92	5.00	4.96
	$p = 3$	5.19	5.31	5.49	5.09	4.98	4.94
	$p = 5$	5.13	5.38	5.93	5.02	4.89	4.28
DGP #2: GARCH	$p = 2$	29.3	29.4	6.12	29.3	4.74	3.24
	$p = 3$	40.3	40.5	8.3	40.1	4.77	2.86
	$p = 5$	51.4	51.8	11.0	51.1	4.17	1.59

6 PT decomposition under cointegration and common cycles

Lettau and Ludvigson (2001) propose a permanent-transitory representation for PVMs using a Gonzalo-Granger decomposition. This decomposition has several drawbacks. However, it corresponds to the Beveridge Nelson (BN hereafter) decomposition if and only if there exist in an n dimensional model (n is the number of series in the VAR), r cointegrating vectors, and exactly $n - r$ serial correlation common feature vectors (see Proietti, 1997; Hecq *et al.* 2000). Hence, the theoretical PVM requirements cannot match the Gonzalo-Granger setup, since the term S_{t-1} must be present in the unpredictable linear combination of the data. In other words, in the PVM there are no linear combinations of ΔY_t and Δy_t alone that are unpredictable, something that is required to use the Gonzalo-Granger decomposition.

Given that PVMs entail weak form common cyclical features, we are able to decompose series into a permanent and a transitory component using the multivariate Beveridge Nelson with common cycles as developed in Hecq *et al.* (2000). Recall that the BN decomposition of $X_t = (Y_t, y_t)'$ in $X_t = \mu_t + \psi_t$, where the trend component denoted μ_t , and ψ_t is a covariance stationary cyclical process, entails the use of the demeaned long-run forecast of $\Delta X_t = (\Delta Y_t, \Delta y_t)'$:

$$\mu_t = X_t + \left\{ \lim_{l \rightarrow \infty} \sum_{i=1}^l \Delta \tilde{X}_{t+i|t} - \mathbb{E}(\Delta X_t) \right\} \quad (31)$$

where $\Delta \tilde{X}_{t+i|t}$ denotes the i th-step ahead forecast. Thus, the trend today represents the value to which the long-term forecast of a series converges to, when we discount its deterministic terms.

Under strong serial-correlation common features (SCCF), Proietti (1997) develops an observable permanent-transitory decomposition of $X_t = (Y_t, y_t)'$ with both common trends and common cycles such that $\beta' \mu_t = 0$ (cointegration), where β is the cointegrating vector, and $\gamma' \psi_t = 0$ (strong SCCF), where γ is the co-feature vector. Components μ_t and ψ_t are derived in Proietti (1997). Now, in the weak-form SCCF, only a part of the cycle is annihilated by γ' . Using the companion form of the VECM, Hecq *et al.* (2000) extend the results in Proietti (1997) and derive an observable decomposition suitable for the weak-form SCCF case, in which $\psi_t = \psi_t^A + \psi_t^B$, with $\gamma' \psi_t^A = 0$, but $\gamma' \psi_t^B \neq 0$ in:

$$X_t = \mu_t + \psi_t^A + \psi_t^B. \quad (32)$$

This can be interpreted either as a decomposition of X_t that includes also WF common cycles or a BN decomposition of $X_t^* = X_t - \psi_t^B$. For more details on the components of (32) see Hecq *et al.* (2000). The decomposition makes X_t to be the sum of three different components: a random walk (martingale) common stochastic trend component (μ_t), a (weak form) common stochastic cycle component (ψ_t^A), and an additional transitory component that is not common to any of the two

series being decomposed (ψ_t^B).

From a different angle, consider a new variable:

$$X_t^* = X_t - \psi_t^B = \mu_t + \psi_t^A. \quad (33)$$

The linear combination $\gamma' X_t^* = \gamma' \mu_t$ is proportional to the common trend, therefore, disregarding deterministic terms, its first difference is unpredictable, i.e., $\gamma' \Delta X_t^*$ is unpredictable. It is worth mentioning that this decomposition is expressed in terms of observables and only involve quantities already available from the VECM form and the estimation of common features and cointegrating vectors.

The fact that we impose the restriction that the trend is a martingale is consistent with the idea that the long-run component of asset prices is captured by a martingale, as put forth by Hansen and Scheinkman (2009). Here, however, our setup is much simpler than theirs, but it still captures the main trust in Beveridge and Nelson that, as in (31), the trend today represents the value to which the adjusted long-term forecast of a series converges to. Any deviations of prices from trend are therefore deviations of prices from fundamentals, which we label here as bubbles. Notice that the concept of a bubble here is intrinsically different from what Campbell and Shiller (1987) and West (1987) have labelled a “rational bubble” and a “speculative bubble,” respectively.

7 Empirical results

We now apply the tools covered in previous sections to two well-known economic issues. On the one hand, these are the relationship between long- and short-term interest rates, and, on the other hand, the relationship between price and dividend. We use the online series maintained and updated by Shiller at <http://www.econ.yale.edu/~shiller/data.htm>. Our investigations are done on those annual data spanning the period 1871-2011 ($T = 141$) for interest rates and on the period 1871-2010 ($T = 140$) for the price-dividend case. For this latter, we divide series by the consumer price index (also from Shiller’s files) in order to obtain real prices and real dividends. Figures 1 and 2 plot the two group of series.

The four variables being $I(1)$ according to usual unit root tests (e.g. ADF), we go next on testing for cointegration and for common cyclical features.

For the interest rates analysis, AIC, SBC and HQ choose $p = 2$ for the VAR in level. For the price-dividend relationship AIC determines $p = 8$, SC $p = 1$ and HQ $p = 2$. For both systems, VAR-residual tests reject the null of no ARCH as well as the null of Normality. In order to choose the lag length of the VAR in the presence of time varying heteroskedasticity, we implement a robust Wald test with the null hypothesis that the last coefficient matrix of a $\text{VAR}(p)$ has only zero coefficients.

Figure 1: Price and dividend series (1871-2010)

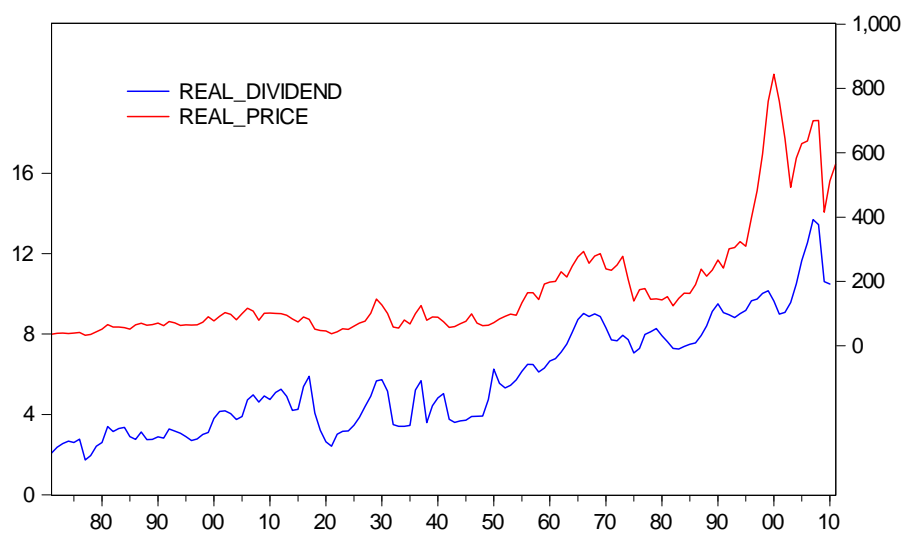


Figure 2: Interest rates series (period 1871-2011)

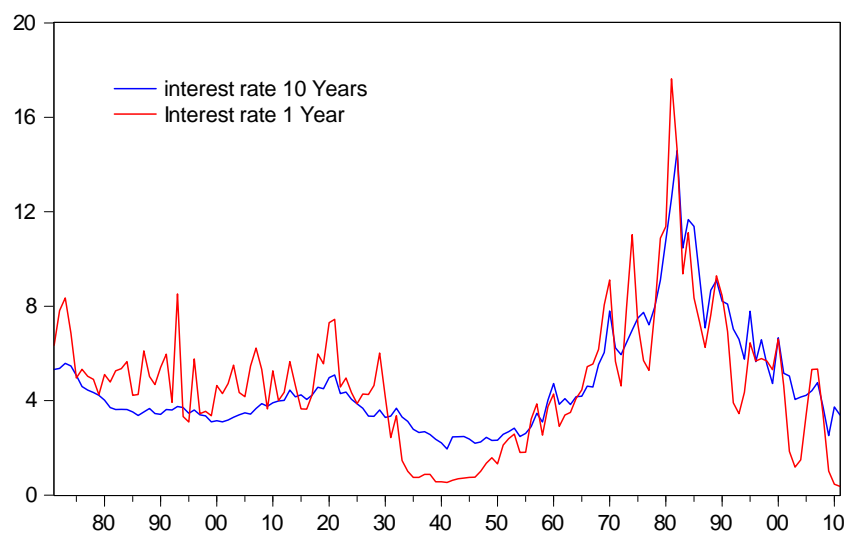


Table 2: Long run properties

i_{lr}/i_{sr}					P/D				
H_0	eigv.	Stat.	cv	p-val	H_0	eigv.	Stat.	cv	p-val
$r = 0$	0.170	28.57	15.49	0.0003	$r = 0$	0.085	13.38	15.49	0.101
$r > 1$	0.018	2.58	3.84	0.1076	$r > 1$	0.007	0.97	3.84	0.322
$i_{lr} = 1.079i_{sr}$ (0.1051)					$P = 67.648D$ (10.195)				
$\alpha_{lr} = -0.0405$ (0.0409)		$\alpha_{sr} = 0.3093$ (0.0895)		$\alpha_P = -0.073$ (0.038)		$\alpha_D = 0.00076$ (0.00052)			

We use a similar strategy that we have already applied in the test ξ_W^{rob} presented above but now with a different set of restrictions. In a bivariate case these can be written as

$$R = I_2 \otimes K,$$

with $K = I_2$ when testing for $p = 1$ (namely the bivariate white noise hypothesis) and $K = [0_{2 \times (p-1)} : I_2]$ when testing for the last lag coefficient matrix when $p > 1$. This test follows a $\chi^2_{(4)}$ under the null. We have investigate on our DGPs this procedure and it emerges that it allows us to determine the correct lag length without any size distortions in the presence of heteroskedastic errors (for our DGP #2). Using this approach, we do not reject the hypothesis of VAR(2) for both economic applications.

As far as long-run properties are concerned, Table 2 shows the Johansen trace test with a constant term for both cases within VAR(2) in levels. Table 2 shows that interest rates are cointegrated with a cointegrating vector very close to $(1 : -1)$ while prices and dividends are only cointegrated with a significance level of about 10%. In this latter case, $\hat{\theta} = 67.648$, the discount factor is $\hat{\delta} = 0.985$ and therefore $\hat{r} = 0.0147$ which is smaller than values found in similar studies. We impose that cointegration exists in both cases, since the 10% level in testing has been used before in this context. Conditional on cointegration we compute spreads using the estimated cointegrated vector for the price/dividend series and imposing a $(1, -1)$ long-run relationship for interest rates.

Given results on long-run properties, we now look at the existence of common cyclical feature relationships. We present the p-values of the six different weak-form common feature test statistics in Table 3. Recall that testing for PSCCF would give numerically identical values. For robustness, we report results for different VAR orders, although we have determined that $p = 2$ using our

Table 3: P-values of the common feature test statistic

		<i>p – values</i>					
	$VAR(p)$	ξ_{LR}	ξ_W	ξ_W^{rob}	J_1	J_2	J_3
i_{lr}/i_{sr}	$p = 2$	0.73	0.73	0.76	0.72	0.75	0.75
	$p = 3$	0.59	0.58	0.72	0.37	0.49	0.45
	$p = 4$	0.54	0.53	0.53	0.11	0.09	0.15
P/D	$p = 2$	0.18	0.18	0.4	0.16	0.46	0.52
	$p = 3$	0.16	0.15	0.25	0.15	0.28	0.20
	$p = 4$	0.36	0.35	0.19	0.35	0.27	0.29

preferred testing procedure in both applications.

In both the long and short-term interest rate case and in the price-dividend analysis the null that there exists at least one weak-form common feature vector is never rejected at usual significance levels. These tests do not reject the necessary rank condition behind the present value theory for both applications, which completes the first step of our proposed testing procedure.

We now further restrict the systems with additional constraints. First we look whether the VECMs are restricted to have a zero first row for lagged dependent series. For the price-dividend case the *p – values* for such an hypothesis is respectively 0.21 and 0.70 using a FIML and a robust GMM approach. If now one adds the restriction that the loading in the price equation is given by the cointegrating vector (i.e., $\hat{r} = 0.0147$) *p – values* are respectively 0.11 and 0.34 for FIML and the robust GMM.

For the long and short-term interest rates case, restricting the dynamics gives *p – values* lower than 0.001 for FIML and 0.04 for robust GMM, hence rejecting the forecasts of the PVM on the value of the coefficients. Hence, for the interest rate analysis we do not further investigate the additional constraints on the coefficients θ, δ and r that are theoretically given by the yield curve (see *inter alia* Johansen and Swensen, 2000, 2004 for a numerical example). Recall that rejecting the PVM can come from two sides: (1) first one can reject the rank deficiency hypothesis and we have seen that this is not the case here; (2) some coefficients might not match their theoretical values. Testing (1) and (2) successively using a common feature framework helps to determine where the problem comes from. Note that in this paper we assume that the series used adequately match the theoretical counterparts in the PVMs.

In order to investigate whether the weak-form common feature relationship is far from the theoretical values for the interest rate case, we obtain the following equation using a pseudo-

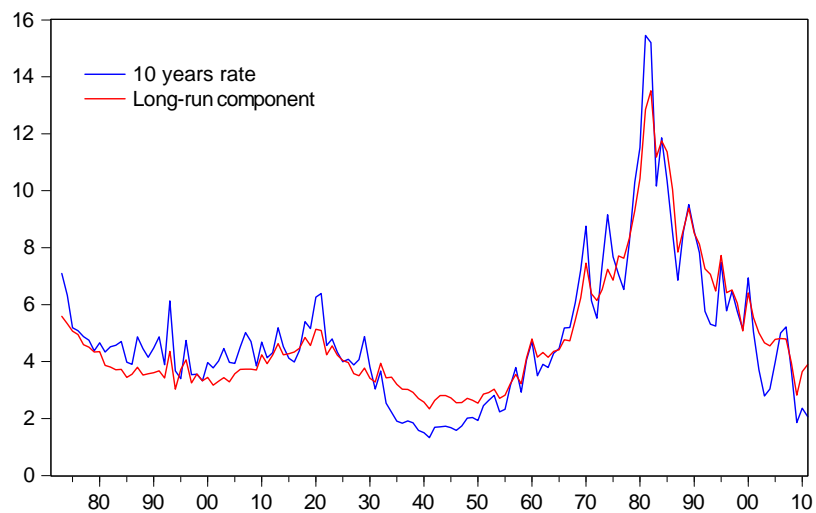
structural model estimated by FIML (standard errors in brackets):

$$\Delta i_{lr,t} = \underset{(0.715)}{1.423} \Delta i_{sr,t} - \underset{(0.193)}{0.463} \hat{S}_{t-1} + \underset{(0.188)}{0.062}.$$

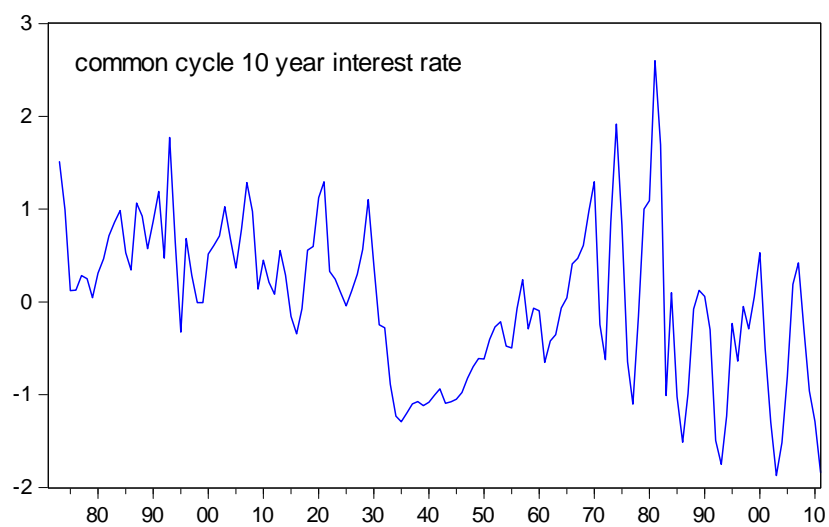
It emerges that the p-value associated with null hypothesis that the coefficient of $\Delta i_{sr,t}$ is equal zero is slightly below 0.05, rejecting (9) at a 5% but not at 10% significance level. Moreover the coefficient being positive and significantly different from -1 , equation (16) which discounts future expected values only, is also rejected. The negative estimated coefficient in front of \hat{S}_{t-1} is also against the predictions of the PVM in both formulations (9) and (16).

Finally, we display for both applications the permanent-transitory decomposition of the series in terms of common trends and common cycles. We only report pictures with constraints coming from cointegration and common features but not with additional restrictions on the values of the coefficients. Figure 3 and Figure 4 plot the main series together with their trend component. Figures 5 and 6 show the cyclical part. For prices and dividends it appears that there is a bubble starting before 2000 in the price series, which we identify as the dotcom bubble. It is also interesting to notice that the recent financial crisis does not seem to have a big deviation from trend. This means that although the situation post 2008 is dramatic for most OECD economies, prices seem to correctly reflect the discounted dividends, something that was missing in the “dotcom bubble.”

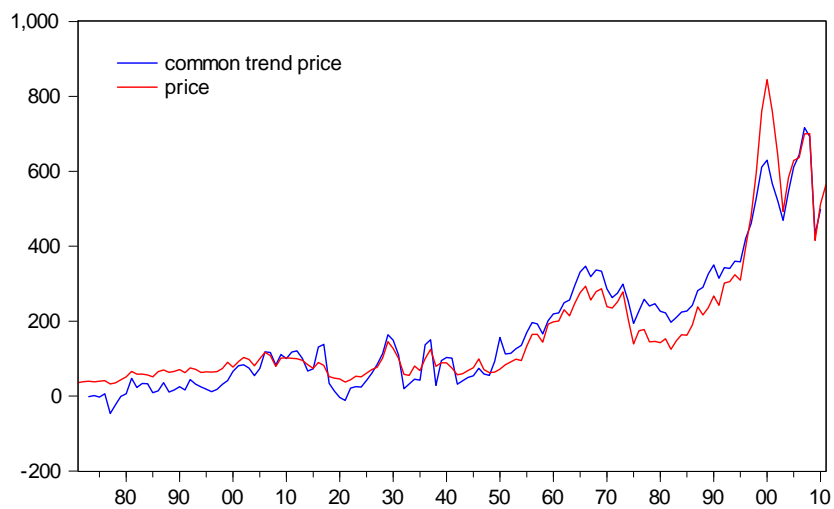
The interest rate cycle is very volatile, especially after the second half of the past century, showing cycles with a duration of about 10 years since the sixties. Notice that, in the last 10 years, the common cyclical component of interest rates is negative, most of the time, reflecting the long-run expectation that they should rise in the future. Indeed, the cyclical component now is in the vicinity of its all time low – almost -200 basis points. The joint analysis of price-dividends and interest rates hints that if there is a bubble for the last global recession it came from interest rates, not from stocks. One of the peculiar features of the last recession is for how long we have observed low levels of short-term interest rates.



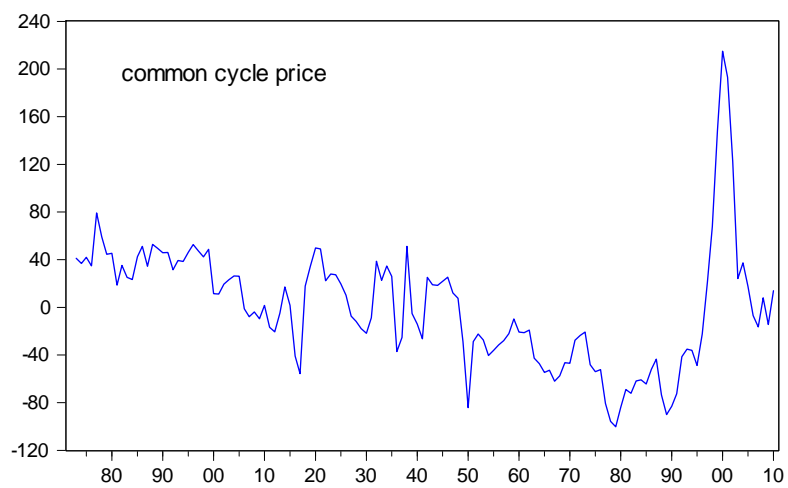
Long term rate and its permanent component



Interest rate common cycle and bubble



Price series and its permanent component



Common cyclical component in prices and bubble

8 Conclusion

The main contribution of this paper is to propose a novel framework for testing PV relationships in economics. Here, we stress that cointegration is simply one side of the restrictions PV relationships impose on the data being tested – it implies that PV equations obey certain transversality conditions. Common-cyclical-feature restrictions form the basis of the other side – they imply the existence of unforecastable errors in the stochastic difference-equation generating PVMs. It is obviously preferable to test for PVMs using an integrated framework where these two types of restrictions are jointly considered. The common-feature *toolkit* allows the investigation of PVMs in multivariate data sets as well as the proposal of new tests for their existence.

Also, in the context of asset pricing, we propose a novel permanent-transitory (PT) decomposition based on Beveridge and Nelson (1981), which focus on extracting the long-run component of asset prices, a key concept in modern financial theory as discussed in Alvarez and Jermann (2005), Hansen and Scheinkman (2009), and Nieuwerburgh, Lustig, and Verdelhan (2010). We advance with respect to standard Beveridge and Nelson (BN) decompositions in which we compute the transitory component directly using the PVM restrictions, which amounts to impose restrictions in the short-run dynamics of the cointegrated VAR to recover the transitory component.

The techniques considered here are applied to two different data sets. The first contains annual long- and short-term interest rates for the US, ranging from 1871 to 2011. The second data set is for the study of the price and dividend relationship on the period 1871-2010. There is one cointegrating relationship and a weak form common feature relationship in both cases, although the analysis of interest-rate data rejects the full set of PVM restrictions. Despite that, the joint analysis of price-dividends and interest rates hints that if there is a bubble for the last global recession it came from interest rates, not from stocks. We are forced to conclude this because the long-run component of asset prices is close to current price, but the same is not true for interest rates. Indeed, the cyclical component of interest rates is now in the vicinity of its all time low – almost -200 basis points.

References

- [1] ALVAREZ, F. AND U. JERMANN (2005), Using asset prices to measure the persistence of the marginal utility of wealth, *Econometrica*, Vol. 73(6), 1977–2016.
- [2] BAILLIE, R. (1989), Econometric tests of rationality and market efficiency, *Econometric Reviews*, 8, 151-186.

- [3] ATHANASOPOULOS, G., GUILLEN, O.T.C., ISSLER, J.V., AND F. VAHID, F. (2011), Model Selection, Estimation and Forecasting in VAR Models with Short-run and Long-run Restrictions, *Journal of Econometrics*, Vol. 164(1), 116-129.
- [4] BEVERIDGE, S. AND C.R. NELSON (1981), A New Approach to Decomposition of Economic Time Series into a Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle', *Journal of Monetary Economics*, 7, 151-174.
- [5] CAMPBELL, J.Y. (1987), Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis, *Econometrica*, vol. 55(6), 1249-73.
- [6] CAMPBELL, J.Y. (1991), A variance decomposition of stock returns, *Economic Journal*, 405, 175-179.
- [7] CAMPBELL, J.Y. AND R.J. SHILLER (1987), Cointegration tests and present value models, *Journal of Political Economy* 95, 1062-1088.
- [8] CAMPBELL, J.Y. AND A. DEATON (1989), Why is Consumption So Smooth?, *The Review of Economic Studies* 56, 357-374.
- [9] CAMPBELL, J.Y., LO, A.W. AND A.C. MACKINLAY (1996), *The Econometrics of Financial Markets*, Princeton University Press.
- [10] CANDELON, B., HECQ, A. AND W. VERSCHOOR (2005), Measuring Common Cyclical Features During Financial Turmoil, *Journal of International Money and Finance*, 24, 1317-1334
- [11] CENTONI, M., CUBADDA, G. AND A. HECQ (2007), Common Shocks, Common Dynamics and the International Business Cycle, *Economic Modelling*, vol.24, 1, 149-166.
- [12] CHOW, G. (1988), Rationale versus adaptive expectations in present value models, Reserach Memorandum 328, Princeton University.
- [13] CHRISTENSEN, T., HURN, S. AND A. PAGAN (2011), Detecting Common Dynamics in Transitory Components, *Journal of Time Series Econometrics*, 2011, vol. 3, issue 1.
- [14] CUBADDA, G. AND A. HECQ (2001), On Non-Contemporaneous Short-Run Comovements, *Economics Letters*, 73, 389-397.
- [15] ENGLE, R. F. AND C.W.J. GRANGER (1987), Co-integration and Error Correction: Representation, Estimation and Testing, *Econometrica*, 55, 251-276.
- [16] ENGLE, R. F. AND S. KOZICKI (1993), Testing for Common Features (with comments), *Journal of Business and Economic Statistics*, 11, 369-395.

- [17] ENGLE, R. F. AND J. V. ISSLER (1995), Estimating Common Sectorial Cycles, *Journal of Monetary Economics*, 35, 83-113.
- [18] ENGSTED, T. (2002), Measures of fit for rational expectations models. *Journal of Economic Surveys* 16, 301-355.
- [19] GONZALO, J. AND C.W.J. GRANGER (1995), Estimation of Common Long-Memory Components in Cointegrated Systems, *Journal of Business and Economics Statistics*, 33, 27-35.
- [20] HAMILTON, J.D. (1994), *Time Series Analysis* (Princeton University Press: Princeton).
- [21] HANSEN, L.P. (1982), Large Sample Properties of Generalized Method of Moment Estimators, *Econometrica*, 50, 1029-1054.
- [22] HANSEN, L. P. AND J. A. SCHEINKMAN (2009), Long-Term Risk: An Operator Approach, *Econometrica*, 77(1), 177 - 234.
- [23] HECQ, A., F. PALM AND J.P. URBAIN (2000), Permanent-Transitory Decomposition in VAR Models with Cointegration and Common Cycles, *Oxford Bulletin of Economics and Statistics*, 62, 511-532.
- [24] HECQ, A., PALM, F.C., J.P. URBAIN (2006), Testing for common cyclical features in VAR models with cointegration, *Journal of Econometrics*, Volume 132(1), 117-141.
- [25] HECQ, A., S. LAURENT AND F.C. PALM (2011), Pure portfolios models, RM maastricht University.
- [26] ISSLER, J.V. AND F. VAHID. F. (2001), Common Cycles and the Importance of Transitory Shocks to Macroeconomic Aggregates, *Journal of Monetary Economics*, vol. 47(3), pp. 449-475.
- [27] JOHANSEN, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models* (Oxford University Press: Oxford).
- [28] JOHANSEN, S. (2000), Modelling of cointegration in the vector autoregressive model, *Economic Modelling*, vol. 17, 359-373.
- [29] JOHANSEN, S. AND A.R. SWENSEN (1999), Testing rational expectations in vector autoregressive models, *Journal of Econometrics*, 93, 73-91.
- [30] JOHANSEN, S., AND A. SWENSEN (2011), On a numerical and graphical technique for evaluating some models involving rational expectations, *Journal of Time Series Econometrics*, 2011, vol. 3, issue 1.

- [31] JOHANSEN, S., AND A. SWENSEN (2004), More on Testing Exact Rational Expectations in Cointegrated Vector Autoregressive Models: Restricted Drift Terms, *Econometrics Journal*, volume 7, pp. 389–397
- [32] LETTAU, M., AND S. C. LUDVIGSON (2001), Consumption, Aggregate Wealth and Expected Stock Returns, *Journal of Finance*, 56(3), 815—849.
- [33] LETTAU, M., AND S. C. LUDVIGSON (2004), Understanding Trend and Cycle in Asset Values: Reevaluating the Wealth Effect on Consumption, *American Economic Review*, 94(1), 276—299.
- [34] LETTAU, M., AND S. C. LUDVIGSON (2010), Measuring and Modeling Variation in the Risk-Return Tradeoff, in *Handbook of Financial Econometrics*, ed. by Y. Ait-Sahalia, and L. P. Hansen, vol. 1, pp. 617—690. Elsevier Science B.V., North Holland, Amsterdam.
- [35] LETTAU, MARTIN, AND SYDNEY C. LUDVIGSON (2011), Shocks and Crashes, NBER Working Paper # 16996.
- [36] PHILLIPS, PETER C B AND HANSEN, BRUCE E. (1990), Statistical Inference in Instrumental Variables Regression with I(1) Processes, *Review of Economic Studies*, 57(1), pp. 99-125.
- [37] PHILLIPS, PETER C B AND LORETAN, MICO (1991), Estimating Long-run Economic Equilibria”, *Review of Economic Studies*, 58, pp. 407-436.
- [38] PROIETTI, T. (1997), Short Run Dynamics in Cointegrated Systems, *Oxford Bulletin of Economics and Statistics*, 59,405-422.
- [39] NEWEY, W. K. AND WEST, K. D., (1987), A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, vol. 55, 703–708.
- [40] RAVIKUMAR, B., RAY, S. AND E. SAVING (2000), Robust Wald tests in Sur systems with adding-up restrictions, *Econometrica*, 68, 715-720.
- [41] VAHID, F. AND R.F. ENGLE (1993), Common Trends and Common Cycles, *Journal of Applied Econometrics*, 8, 341-360.
- [42] VAHID, F. AND R.F. ENGLE (1997), Codependent Cycles, *Journal of Econometrics*, 80, 199-221.
- [43] VAHID, F., J.V. ISSLER (2002), The importance of common cyclical features in VAR analysis: A Monte Carlo study, *Journal of Econometrics* 109, 341–363.

- [44] WEST, KENNETH D. (1987), A Specification Test for Speculative Bubbles, *The Quarterly Journal of Economics*, Vol. CII, No. 3, pp. 553-580.